

## **Paper 2**

**Title:**Power Analysis to Detect the Effects of a Continuous Moderator in 2-Level Simple Cluster Random Assignment Experiments

### **Authors and Affiliations:**

Nianbo Dong  
College of Education  
University of Missouri

**Background / Context:**

For intervention studies involving binary treatment variables, procedures for power analysis have been worked out and computerized estimation tools are generally available. Of greatest importance for educational research, such procedures and tools are available for the kinds of complex multilevel designs that are often required for testing education interventions, for example, Bloom (1995, 2005, 2006), Hedges & Rhoads (2010), Konstantopoulos (2008a, 2008b, 2009, 2010, 2012), Murray (1998), Raudenbush (1997), Raudenbush & Liu (2000), Raudenbush, Martinez, & Spybrook (2007), and Schochet (2008). Among the excellent computer programs available for conducting power analysis for cluster randomized experiments are Optimal Design (Raudenbush, Spybrook, Congdon, Liu, & Martinez, 2011), CRT-Power (Borenstein & Hedges, 2012), and PowerUp! (Dong & Maynard, 2013).

However, there are relationships other than the main effects of binary treatment variables that interest education researchers. For example, researchers may wish to determine if some classroom practice described by a continuous variable, such as the number of disruptive incidents or the amount of time on task, is related to student academic achievement when other influential classroom characteristics are statistically controlled. Or, within the context of an experimental study, researchers may wish to determine whether the effects of the intervention differ conditional on moderator variables such as pretest, ethnicity, school climate, or the fidelity of implementation. Power analysis for relationships between continuous predictors and dependent variables in multivariate, multilevel models cannot be accomplished with the procedures and estimation tools that have been developed for intervention studies. Moreover, power for moderator analyses in multilevel intervention studies cannot be estimated using the procedures and tools for estimating the power of main effects. Compared to power analysis for main effects in cluster randomized experiments, there is less support for power analysis for relationships involving continuous predictors or the interactions of moderator variables with treatment effects. The only research we have identified on power analysis for relationships of continuous predictors with outcomes in a multivariate multilevel model is Snijders & Bosker's (1993), which was restricted to a two-level HLM, and did not result in any computational tools for use by researchers. For moderator relationships in experimental studies, Bloom (2005) and Spybrook (in press) have presented procedures for conducting power analysis for binary moderators in two- to four-level cluster randomized experiments, but have not extended those procedures to include continuous moderator variables. Most recently, Mathieu, Aguinis, Culpepper, & Chen (2012) conducted a comprehensive Monte Carlo simulation to estimate the statistical power to detect cross-level interaction effects. However, Mathieu et al (2012) only studied two-level analysis without including covariates, and did not provide closed form formulas to estimate the statistical power, minimum detectable effect size, or minimum required sample size to detect meaningful effects.

**Purpose / Objective / Research Question / Focus of Study:**

The purpose of this study is to: (1) develop the statistical formulations for calculating statistical power, minimum detectable effect size (MDES) and its confidence interval, and minimum required sample size to detect the effects of a continuous moderator variable at level 1 or level 2 in two-level simple cluster random assignment designs, and (2) operationalize these formulas in the enhanced version of PowerUp! (Dong & Maynard, 2013) to create spreadsheets for calculating MDES, etc.

**Significance / Novelty of study:**

Educational researchers have interests in the effects of continuous moderators in cluster-randomized experiments. Statistical power analysis is appropriate in the planning stages to help researchers design studies with sufficient power to detect such relationships when they are large enough to have practical or theoretical significance. However, currently there is no tool available for researchers to conduct such power analysis. This study will provide a tool for power analysis to detect the continuous moderator effect in two-level simple cluster random assignment designs.

## **Research Design:**

### *Framework*

We use the framework of the minimum detectable effect (MDE) popularized by Bloom (1995, 2005, 2006; Murray, 1998). Although this framework was originally developed for estimating the power for detecting the relationship between a binary treatment variable and an outcome variable, it can be extended to other relationships with dependent variables in multilevel models. MDE measured in raw scale units can be expressed as  $MDE(b_0) = M_v SE(b_0)$ , where  $b_0$  is the effect (unstandardized coefficient) of the focal predictor,  $SE(b_0)$  is the standard error of that effect, and  $M_v$  is a multiplier that carries information about the selected alpha level, statistical power target, and degrees of freedom for the significance test. Specifically,  $M_v$  is the sum of two  $t$ -statistics. For one-tailed tests,  $M_v = t_\alpha + t_{1-\beta}$  with  $v$  degrees of freedom (a function of sample size and number of covariates), and for two-tailed tests,  $M_v = t_{\alpha/2} + t_{1-\beta}$ . MDE, in turn, is the minimum effect ( $b_0$  in raw scale units) that can be detected at the  $\alpha$  level with probability (statistical power)  $1 - \beta$ .

When the focal predictor is a continuous variable, defining the effect size as the standardized mean difference is not appropriate. For the derivations to be developed in this project, we will use the standardized regression coefficient ( $\beta_0$ ) for the predictor of interest as an effect size. This can be estimated using HLM by first standardizing the outcome variable and the predictors, i.e.,  $N(0,1)$ , as the effect size metric. Note that this standardized coefficient is equal to the Pearson correlation ( $r$ ) of the predictor and the outcome when there is only one predictor in the model and it is a semipartial correlation coefficient when there are two or more predictors.

An alternate representation of the effect for the designs and analyses of interest here is the correlation between the predictor and the dependent variable. In multilevel analysis, however, there are multiple possible expressions of this correlation. For instance, in a two-level analysis with only one level-2 predictor, the correlation of the predictor and the individual (level 1) outcome values (which is the same as the standardized regression coefficient), and the correlation of the predictor and the cluster means on the outcome variable at level 2 are different, but both represent the association of the predictor and the outcome and both can serve as an effect size metric. For some derivations, it is convenient to represent the effect in terms of the correlation at the level of the focal predictor (adjusted as needed for additional covariates in the model), but this can be easily converted to the corresponding standardized regression coefficient.

Similarly, in the analysis of the effect of the moderator in multilevel experiments, the standardized coefficient of the interaction term for the treatment variable and the moderator is the effect size of primary interest. This coefficient can also be expressed in correlational terms as a function of the separate correlations between the predictor and the outcome for the control and treatment groups. The standardized coefficient and the difference between the treatment and control correlations (adjusted for other covariates as needed) are alternate representations of the

interaction effect and one can be converted to the other.

As is typical in multilevel power analysis, we assume that the data are balanced such that each cluster has the same number of observations ( $n$ ). We also assume that each cluster has the same values of the predictors and the same sampling variance for the coefficient of the predictor. Under these assumptions, the unique, minimum-variance, unbiased estimator of the coefficient ( $b_0$ ) of the focal continuous predictor would be the OLS (ordinary least square) regression estimator (Raudenbush & Bryk, 2002, p.43). Hence, we will use the OLS estimators to: (1) derive formulas for the estimate of the coefficient ( $b_0$ ) of the focal continuous predictor or the interaction term for a moderator relationship expressed by the correlation ( $r$ ) or the standardized coefficient ( $\beta_0$ ), and (2) derive formulas for the estimate of the standard error ( $SE(b_0)$ ) of the coefficient expressed as the correlation ( $r$ ) or the standardized coefficient ( $\beta_0$ ). Furthermore, the minimum detectable effect should satisfy  $M_v = MDE(b_0) / SE(b_0)$ . The multiplier can be calculated based on the desired  $\alpha$  level, statistical power  $1 - \beta$ , sample size, and the number of covariates. Hence, the correlation ( $r$ ) or the standardized coefficient ( $\beta_0$ ) can be expressed as a function of the multiplier,  $M_v$ , and the sample size, etc.

We illustrate the procedures (but omit some details due to page limitation) for the derivations of MDES with two analytic examples. The first example is a basic two-level hierarchical linear model (HLM) that includes one level-2 focal continuous predictor,  $W$ , and no covariates. The second example is a two-level cluster randomized design with the sample equally divided between the treatment and control groups and a continuous moderator variable at level 2. The procedures illustrated for these two examples will be extended to the other design and analysis models for which we propose to develop power analysis formulations with adaptations that take the greater complexity of those designs into account.

#### *Two-level HLM with a level-2 continuous predictor*

The HLM including one level-2 continuous predictor,  $W_j$ , is:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij}, \quad r_{ij} \sim N(0, \sigma^2)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j}, \quad u_{0j} \sim N(0, \tau_w^2)$$

Based on results from Raudenbush and Bryk (2002, pp. 39-41), we can derive MDES in terms of the correlation of the sample cluster mean,  $\bar{Y}_j$ , and the level-2 predictor,  $W_j$ , for a given level-2 sample size,  $J$ , desired  $\alpha$  level, and statistical power ( $1 - \beta$ ) as:

$$(1) \quad MDES(|r_{\bar{Y}, W}|) = \sqrt{\frac{M_v^2}{M_v^2 + J - 2}}, \quad \text{where, } M_v = t_\alpha + t_{1-\beta} \text{ for one-tailed tests with } v \text{ degrees of freedom (} v = J - 2 \text{ when there is only one predictor), and } M_v = t_{\alpha/2} + t_{1-\beta} \text{ for two-tailed tests.}$$

The MDES in terms of the standardized coefficient can be derived:

$$(2) \quad MDES(|\gamma_{01}|) = \sqrt{\frac{M_v^2}{M_v^2 + J - 2}} \sqrt{\rho + (1 - \rho)/n}, \quad \text{where } \rho \text{ is the unconditional intra-class correlation (ICC).}$$

#### *Two-level cluster randomized design with a continuous moderator at level 2*

The full HLM for this example, including one treatment variable,  $T_j$ , and one level-2 moderator,  $W_j$ , is:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + r_{ij}, \quad r_{ij} \sim N(0, \sigma^2)$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}W_j + \gamma_{02}T_j + \gamma_{03}(W_j \times T_j) + u_{0j}, \quad u_{0j} \sim N(0, \tau_{W,T}^2)$$

The interest for moderator analysis is whether the parameter,  $\gamma_{03}$ , which indicates the relationship between the treatment effect and the moderator, is statistically significant.

Based on the previous results for a continuous level-2 predictor (Expressions 1 & 2), the minimal detectable effect sizes in terms of the incremental Cohen's  $f$  and the correlation ( $r_{\bar{Y},W}^t$ ) of  $\bar{Y}_j$  and  $W_j$  for the treatment group can be calculated from Expressions 3 and 4:

$$(3) \quad f_t - f_c = M_v \sqrt{\frac{4}{J-4}}, \text{ and}$$

$$(4) \quad r_{\bar{Y},W}^t = \sqrt{\left( M_v \sqrt{\frac{4}{J-4}} + \sqrt{\frac{(r_{\bar{Y},W}^c)^2}{1-(r_{\bar{Y},W}^c)^2}} \right)^2 / \left( 1 + \left( M_v \sqrt{\frac{4}{J-4}} + \sqrt{\frac{(r_{\bar{Y},W}^c)^2}{1-(r_{\bar{Y},W}^c)^2}} \right)^2 \right)},$$

where  $M_v = t_\alpha + t_{1-\beta}$  for one-tailed tests with  $v$  degrees of freedom ( $v = J - 4$  when HLM includes the treatment variable, moderator, and the interaction term for the treatment and

moderator), and  $M_v = t_{\alpha/2} + t_{1-\beta}$  for two-tailed tests. Let Cohen's  $f$ ,  $f_t = \sqrt{\frac{(r_{\bar{Y},W}^t)^2}{1-(r_{\bar{Y},W}^t)^2}}$  and

$$f_c = \sqrt{\frac{(r_{\bar{Y},W}^c)^2}{1-(r_{\bar{Y},W}^c)^2}}, \text{ i.e., } r_{\bar{Y},W}^t = \sqrt{\frac{f_t^2}{1+f_t^2}} \text{ and } r_{\bar{Y},W}^c = \sqrt{\frac{f_c^2}{1+f_c^2}}.$$

Note that the original Cohen's  $f$  is one metric of effect size for OLS regression. We have two Cohen's  $f$ s for the control and treatment groups. It is the incremental Cohen's  $f$  (that is,  $f_t - f_c$ ) that represents the effect size of the moderator effect, i.e., the effect difference of the predictor between the treatment and control groups. The above results suggest that the incremental Cohen's  $f$  due to the intervention, and the sample size ( $J$ ) must be large enough to satisfy Expression 4 to detect a significant differential effect ( $\gamma_{03}$ ) at the  $\alpha$  level with statistical power of  $1 - \beta$ .

### Results and Conclusions:

This abstract only shows the preliminary results of the MDES for detecting a level-2 continuous moderator effect in two-level cluster randomized experiments. The formulas for statistical power and the minimum required sample sizes will be derived accordingly. Furthermore, the MDES etc. for a level-1 continuous moderator and with covariates in two-level cluster randomized experiments will be derived. All these formulas can be operated in Microsoft Excel in the enhanced version of PowerUp! to help researchers with designing moderation analysis in multilevel experiments.

## Appendix A. References

- Bloom, H. S. (1995). Minimum detectable effects: A simple way to report the statistical power of experimental designs. *Evaluation Review*, 19 (5), 547-556
- Bloom, H. S. (2005). Randomizing groups to evaluate place-based programs. In Howard S. Bloom (editor), *Learning more from social experiments: Evolving analytic approaches*, 115-172, New York: Russell Sage Foundation.
- Bloom, H. S. (2006). The core analytics of randomized experiments for social research. MDRC Working papers on research methodology. Available online at: <http://www.mdrc.org/publications/437/full.pdf>
- Borenstein, M. & Hedges, L. V. (2012). CRT-Power – Power Analysis for Cluster-Randomized and Multi-Site Studies [Computer software]. Englewood, NJ: Biostat.
- Dong, N., & Maynard, R. A. (2013). PowerUp!: A tool for calculating minimum detectable effect sizes and minimum required sample sizes for experimental and quasi-experimental design studies. *Journal of Research on Educational Effectiveness*, 6(1), 24-67. doi: 10.1080/19345747.2012.673143. PowerUp! is available at: <http://web.missouri.edu/~dongn/>
- Hedges, L. V. & Rhoads, C. (2010). *Statistical power analysis in education research* (NCSE 2010-3006). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. This report is available on the IES website at <http://ies.ed.gov/ncser/>.
- Konstantopoulos, S. (2008a). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, 1, 66-88.
- Konstantopoulos, S. (2008b). The power of the test for treatment effects in three-level block randomized designs. *Journal of Research on Educational Effectiveness*, 1, 265-288.
- Konstantopoulos, S. (2009). Using power tables to compute statistical power in multilevel experimental designs. *Practical Assessment, Research & Evaluation*, 14(10). Retrieved April 21, 2010: : <http://pareonline.net/getvn.asp?v=14&n=10>.
- Konstantopoulos, S. (2010). Power analysis in two-level unbalanced designs. *The Journal of Experimental Education*, 78, 291-317. doi: 10.1080/00220970903292876.
- Konstantopoulos, S. (2012). The impact of covariates on statistical power in cluster Randomized designs: Which level matters more? *Multivariate Behavioral Research*, 47, 392-420.
- Mathieu, J .E., Aguinis, H., Culpepper, S. A., and Chen, G. (2012). Understanding and

- estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, 97 (5), 951-966.
- Murray, D. (1998). *Design and analysis of group-randomized trials*. Oxford: Oxford University Press.
- Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods*, 2(2), 173-185.
- Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Raudenbush, S. W. & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, 5(2), 199-213.
- Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for Improving Precision in Group-Randomized Experiments. *Educational Evaluation and Policy Analysis*. 29 (1): 5-29.
- Raudenbush, S. W., Spybrook, J., Congdon, R., Liu, X., & Martinez, A. (2011). Optimal Design Software for Multi-level and Longitudinal Research (Version 2.01) [Computer software]. [www.wtgrantfoundation.org](http://www.wtgrantfoundation.org).
- Schochet, P. Z. (2008). Statistical power for randomized assignment evaluation of education programs. *Journal of Educational and Behavioral Statistics*. 33 (1), 62-87.
- Snijders, T. A. B. & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics*, 18, 237-259.
- Spybrook, J. (in press). Detecting intervention effects across context: An examination of the power of cluster randomized trials. *Journal of Experimental Education*.